

Cargo and Semantic MediaWiki

Yaron Koren
SMWCon Fall 2015
Barcelona, Spain
October 29, 2015

Semantic MediaWiki: overly complicated?

In most setups, SMW has:

- A table structure (e.g., MySQL)
- representing a graph structure (triplestore)
- representing a table structure (infobox data)

Most of the time, yes, it is more complicated than it needs to be.

Cargo

- A MediaWiki extension created and mostly written by me
- First released in January 2015
- Does **storage and querying of data**
- Intended to be an alternative to the main SMW “ecosystem”: SMW + SRF, SM, Maps, SD, SCQ, SIO, + the DataValues extensions

How does it work?

- Uses a “table” structure, not a “graph” structure
- For each infobox template/category, a separate DB table is created
- Each template *call* adds one row to that template's table
- Cargo query language is a thin (and safe!) wrapper around SQL

Cargo parser functions

- **#cargo_declare** – declares a DB table (similar to SQL “CREATE TABLE”)
- **#cargo_store** – stores one row of a table (similar to SQL “INSERT INTO”)
- **#cargo_query** – runs a query (similar to SQL “SELECT”)
- **#cargo_compound_query** – displays multiple queries at once

Sample storage calls

In `<noinclude>` part of template:

```
{{#cargo_declare:_table=Countries|  
Date_founded=Date|Official_languages=List  
(,) of Page}}
```

(creates table)

In `<includeonly>` part of template:

```
{{#cargo_store:_table=Countries|  
Date_founded={{Date founded|}}} |  
Official_languages={{Official  
languages|}}} }}
```

(creates table row)

Sample query 1

Get all cities in this country:

```
{ {#cargo_query:  
tables=Cities  
|fields=_pageName=City,Date_founded  
|where=Country="{ { PAGENAME} }"  
} }
```

Sample query 2

Get all cities in Europe:

```
{ {#cargo_query:  
tables=Cities,Countries  
|join on=Cities.Country =  
Countries._pageName  
|fields=Cities._pageName,  
Countries._pageName  
|where=Countries.Continent = "Europe"  
|format=dynamic table  
} }
```


Cargo query formats

Essentially the same set that SMW/SRF/SM provide:

- Lists, bulleted lists, tables, etc.
- “template” format
- Calendars, timelines
- Maps
- Bar charts
- Exhibit format (similar to “filtered”)

Advantages of Cargo over Semantic MediaWiki

#1: Easier to install

- One extension (Cargo) replaces around 15
- Much less need for solutions like Composer or Semantic Bundle

#2: Easier to understand

- A table structure for data (Cargo's approach) is more familiar than a graph structure (SMW's approach)
- “Everything in the template” is a simpler concept than dealing with properties

#3: Easier to set up data structures

- No property pages (these can easily be 95% of the data structure)
- No “subobjects” (or “internal objects”) - all data is stored the same way
- No need to set up drilldown filters

#4: Easier to set up queries

- Cargo's querying system is based on SQL, a widely-known technology – less learning time for most people

#5: More powerful querying

Queries possible in Cargo but not in SMW include:

- “Joined” data, e.g. “? Has country.Has flag”
 - In Cargo: join two (or more) tables
- Get all pages with a blank value for a field
 - In Cargo: `where=FieldName == ' ' OR FieldName IS NULL`
- Complex combinations of AND, OR and NOT

#6: Faster querying

- Because of Cargo's simpler d0ata structure, querying is faster
- One test was done (on discoursedb.org), using equivalent data sets: Cargo's query ran 50% faster (2/3 of the time) than the equivalent SMW query

#7: Easier querying/analysis from outside the wiki

- In SMW: use the API (and cycle through if more results than the query limit)
- In Cargo: use the API, or – if you have access – **query the DB tables directly**

#8: Easier to develop and maintain the software

Not counting 3rd-party JS libraries:

- SMW: **100,000** lines of code
- SRF: **20,000** lines of code
- SMW's “library” extensions (DataValues, etc.): **50,000** lines of code
- Cargo: **10,000** lines of code

Advantages of Semantic MediaWiki over Cargo

#1: Storage of arbitrary data

SMW lets you just add random properties to any page.

#2: Hierarchical querying of categories

You can query on “World leaders” and also get Presidents, Kings, etc. to show up in the list.

#3: Querying of page metadata

In SMW (plus Semantic Extra Special Properties), you can get “Date created”, etc. for each page.

This may be coming to Cargo!

#4 RDF storage and export

These could be added to Cargo, in theory.

#5: Notifications via Semantic Watchlist

Currently no good notification system exists for non-SMW wikis.

Am I destroying this community?

I don't think so.

SMW is more than properties and #ask queries.

“Semantic MediaWiki” has come to represent an entire text-based approach to data that is *still unique*.

We are also the energy behind, and the voice of, “enterprise MediaWiki”.

None of this will change even if everyone switches to Cargo.

What about SMWCon?

I hope it will continue!

Perhaps under a different name?

Perhaps with a broader focus?

Questions/comments/complaints

